

Age-Neutrality of the Trait Facets Proposed for Personality Disorders in DSM-5: A DIFAS Analysis of the PID-5

Joke Van den Broeck · Leen Bastiaansen · Gina Rossi ·
Eva Dierckx · Barbara De Clercq

Published online: 27 June 2013
© Springer Science+Business Media New York 2013

Abstract An age-neutral measurement system is one of the basic conditions for an accurate personality assessment across the lifespan, both longitudinally and cross-sectionally. In this study the age-neutrality of the *Personality Inventory for DSM-5* PID-5; Krueger et al. (Psychological Medicine 42, 1879–1890, 2012) was investigated. Potential Differential Item Functioning (DIF) was examined for the 25 trait facets in older versus younger adults. Overall, 33 items displayed large DIF, according to the adjusted Bonferroni corrected cutoffs (Mantel Chi-square, Liu-Agresti Cumulative Common Log-Odds Ratio [L-A LOR], and Cox's Noncentrality Parameter Estimator [Cox's B]). In a next step, the implications of the item level DIF across age groups was investigated on scale (i.e., facet) level. These Differential Test Functioning (DTF) analyses revealed large DTF for four of the 25 PID-5 facets (i.e., Withdrawal, Attention Seeking, Rigid Perfectionism and Unusual Beliefs). Current initial results show that most PID-5 traits are measured equally well across age, however, further research is needed to further refine this instrument and make it entirely age-neutral.

Keywords Assessment · Personality disorders · Age-neutrality · Differential item functioning · DSM-5 · PID-5

Introduction

The *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5; American Psychiatric Association [APA] 2013) retained the categorical DSM-IV (APA 1994) approach to personality disorders, describing ten different personality disorder categories. Many prominent researchers in the field advocated a dimensional approach (e. g., Widiger and Trull 2007) and during the development process of DSM-5, the DSM-5 Personality and Personality Disorders Work Group (<http://www.dsm5.org/MeetUs/Pages/PersonalityDisorders.aspx>) proposed a hybrid dimensional-categorical model for personality and personality disorder assessment and diagnosis in which six specific personality disorder types are defined by two fundamental criteria, being impairments in personality functioning and the presence of pathological personality traits. Regarding the latter, a multidimensional maladaptive personality trait system has been developed, in order to represent individual differences in personality disorder expression (Krueger et al. 2012; Wright et al. 2012). In this model, 25 primary traits are organized by five higher-order dimensions: Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism. The *Personality Inventory for DSM-5* (PID-5; Krueger et al. 2012) was developed to operationalize these DSM-5 traits. The above mentioned alternative dimensional proposal is included in a separate chapter of DSM-5 (APA 2013), “Results”, and aims to encourage further research on how this new methodology can be used to assess and diagnose personality disorders. Unfortunately, the Work Group's dimensional focus detracted attention from another important issue, namely the suitability of the criteria for measuring personality in later life (Tackett et al. 2009; Oltmanns and Balsis 2011). This is regrettable, since “even if a dimensional shift is made, there will be continued psychometric and conceptual problems if the criteria do not closely consider the presentation of personality in later life” (Tackett et al. 2009, p. 14). It is commonly assumed that the current

J. Van den Broeck · L. Bastiaansen · G. Rossi · E. Dierckx
Department of Clinical and Lifespan Psychology,
Vrije Universiteit Brussel, Brussels, Belgium

B. De Clercq
Department of Developmental, Personality and Social Psychology,
Ghent University, Ghent, Belgium

J. Van den Broeck (✉)
Faculty of Psychology and Educational Sciences,
Department of Clinical and Lifespan Psychology,
Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
e-mail: Joke.Van.Den.Broeck@vub.ac.be

DSM-IV criteria for personality disorders are not adequately attuned to the living situations and experiences of older people (Agronin and Maletta 2000; Segal et al. 2000). Based on item response theory analyses in a large, cross-sectional study of 37,000 participants, Balsis et al. (2007) concluded that 29 % of the DSM-IV Axis II criteria lack face validity in older age groups, hence possibly leading to over- or underdiagnosis of personality pathology in old age populations. Unfortunately, the presentation of later life was not explicitly considered in the construction of the PID-5 either. Yet an age-neutral measurement system is one of the basic conditions for an accurate personality assessment across the lifespan, both longitudinally and cross-sectionally.

Since the PID-5's age-neutrality has, to our knowledge, not been empirically investigated thus far, we set out to detect possible Differential Item Functioning (DIF) in older versus younger adults. An item is said to exhibit DIF when younger and older adults with a similar position on the underlying trait of interest do not have the same probability of endorsing that item (Edwards and Edelen 2009). If DIF occurs, the assumption of measurement invariance is violated, leading to possibly flawed interpretations of observed between-group differences (Millsap 2011). DIF analyses can be done in both a Classical Test Theory (CTT) and an Item Response Theory (IRT) framework. The current analyses were conducted using an odds ratio approach (CTT), which is, in contrast to the IRT approach, not hampered by requirements of model fit and large sample sizes, and can be conducted using the easily accessible DIFAS program (Penfield 2005). In addition, it has been shown that the cumulative common log-odds ratio is very similar to the IRT based approaches (Penfield 2010).

Method

Participants and Procedure

A total of 464 participants were included in the current study, subdivided into a younger and an older sample. The younger sample consisted of 288 undergraduate psychology students with ages ranging from 17 to 40 ($M=21.05$, $SD=3.70$, 27 % male). Participants in the older sample were 176 Dutch-speaking community-dwelling adults recruited by undergraduate psychology students. Students were requested to recruit at least one person aged 60 or older. No other specifications or conditions were provided. As a return for participation, the students received course credits. Participants' age in the older sample ranged between 61 and 99 years ($M=72.73$; $SD=6.09$), with 40 % male participants. All participants volunteered their participation and provided a written informed consent.

Measure

The Dutch authorized translation of the *Personality Inventory for DSM-5* (PID-5; Krueger et al. 2012; De Clercq et al. 2011), a 220-item self-report questionnaire, was used to measure the proposed DSM-5 traits. Items are rated on a 4-point Likert format scale, ranging from very false or often false to often true or very true. The PID-5 has 25 primary lower-order scales or facets that load onto five higher-order personality pathology dimensions (Negative Affectivity, Detachment, Antagonism, Disinhibition, and Psychoticism). Lower-order scale internal consistencies (Cronbach's alpha) ranged from 0.68 (Suspiciousness) to 0.95 (Eccentricity) in the current younger sample, and from 0.25 (Suspiciousness) to 0.91 (Eccentricity) in the older sample ($Mdn=.82$ in both samples). These values are largely in line with data reported in previous research (e.g., Wright et al. 2012; Hopwood et al. 2012), however the internal consistency of Suspiciousness is remarkably low in the current older sample (see Table 1). Closer examination of the low alpha value for Suspiciousness (i.e., 0.25) revealed that this is mainly due to the two reversed scored items in this scale (i.e., i131 'people are basically trustworthy', and i177 'I rarely feel that people I know are trying to take advantage of me'). After deleting these two items from the scale, the alpha value increased to 0.54.

Statistical Analyses

Between-group differences between the younger and older age group on the PID-5 facets were computed with t-tests for independent samples. Cohen's d was used as a measure of effect size (Cohen 1988), with $d \geq .20$ indicating a small effect, $d \geq .50$ a medium effect, and $d \geq .80$ a large effect).

Next, it was investigated whether there are PID-5 items that measure the personality trait of interest differently in younger versus older adults, after controlling for the overall level of underlying trait between both age groups. To detect possible DIF as a function of age, the Mantel Chi-square (Mazor et al. 1992), the Liu-Agresti Cumulative Common Log-Odds Ratio (L-A LOR; Liu and Agresti 1996), and the Cox's Noncentrality Parameter Estimator (Cox's B; Camilli and Congdon 1999) were used for polytomous items, using the DIFAS 5.0 software program (Penfield 2007). According to these three approaches, individuals are first divided across different "bands" or "strata", based on their total scale scores (measuring the personality trait of interest). All individuals within the same stratum are considered equated, so that any difference between individuals in the younger and older age group are considered indicative of DIF (Baer et al. 2011). The Mantel chi-square statistic is based on a group (2) $\times$ response option (4) contingency table, distributed as a chi-square with 1° of freedom. The higher the chi-square value,

Table 1 Descriptive statistics for the PID-5 primary traits for the younger ($n=288$) and older ($n=176$) sample

Facet	Cronbach alpha		Raw means (SD)		Effect size Cohen's d
	Young	Old	Young	Old	
Anhedonia	.83	.74	4.97 (3.60)	5.73 (3.88)	-.20*
Anxiousness	.88	.85	11.05 (5.59)	8.44 (5.61)	.47***
Depressivity	.89	.88	7.25 (6.01)	6.18 (6.45)	ns
Emotional Lability	.88	.85	9.24 (4.94)	7.17 (4.86)	.42***
Hostility	.81	.78	10.86 (4.88)	7.76 (5.13)	.62***
Perseveration	.78	.74	8.56 (4.22)	7.32 (4.49)	.28**
Rigid Perfectionism	.87	.85	9.93 (5.77)	11.02 (6.10)	ns
Separation Insecurity	.74	.72	9.02 (3.87)	8.23 (4.19)	.20*
Submissiveness	.76	.73	4.18 (2.24)	3.94 (2.76)	ns
Suspiciousness	.68	.24	7.02 (3.15)	7.39 (2.64)	ns
Withdrawal	.87	.87	4.66 (4.52)	6.34 (5.67)	-.33***
Attention Seeking	.86	.86	8.13 (4.39)	5.34 (4.83)	.60***
Callousness	.81	.77	5.58 (4.34)	5.60 (5.24)	ns
Deceitfulness	.84	.84	7.20 (4.61)	4.74 (4.97)	.51***
Grandiosity	.72	.83	2.83 (2.43)	2.57 (3.26)	ns
Manipulativeness	.83	.82	5.07 (3.13)	2.91 (3.12)	.69***
Intimacy Avoidance	.81	.68	2.15 (2.74)	5.79 (3.67)	-1.12***
Restricted Affectivity	.82	.70	5.69 (3.97)	6.01 (3.62)	ns
Distractibility	.89	.84	10.33 (5.42)	7.10 (5.29)	.60***
Eccentricity	.95	.91	9.05 (8.01)	6.33 (7.00)	.36***
Perceptual Disregulation	.79	.86	5.13 (4.51)	4.97 (5.53)	ns
Risk Taking	.89	.74	18.64 (6.87)	13.27 (5.83)	.84***
Unusual Beliefs	.80	.81	2.88 (3.42)	3.57 (4.03)	ns
Impulsivity	.80	.72	6.57 (3.26)	5.06 (3.44)	.45***
Irresponsibility	.71	.71	4.91 (3.05)	3.08 (3.13)	.59***

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

the higher the probability the item displays DIF. In line, the L-A LOR considers the log odds ratio of one group endorsing a response option relative to another. The Cox's B statistic is similar to the Mantel-Haenszel statistic but uses the hypergeometric mean. For both the L-A LOR and Cox's B statistic negative values indicate DIF against the reference group (younger adults), and positive values indicate DIF against the focal group (older adults). For example, an item revealing a negative L-A LOR and Cox's B value is more readily endorsed by older adults, despite the fact that both groups were matched on the underlying personality trait of interest (e.g., Anxiousness). The following cut-off criteria are available to flag items with large DIF: $|L-A LOR| > .64$ (Penfield 2007), and $|Cox's B| > .40$ (Camilli and Congdon 1999). Critical values for the Mantel Chi-square statistic are 3.84 for a Type I error rate of 0.05, and 6.63 for a Type I error rate of 0.01 (Penfield 2007). However, to reduce the Type I error of incorrectly flag items as problematic, we used a stringent Bonferroni corrected critical chi-square value, ranging from 8.28 to 10.83 (depending on the number of

items per scale; for example for Anhedonia the Bonferroni correction was applied across all 8 items by three test statistics $[0.05/8*3]$, leading to a critical chi-square value of 9.55). Similar stringent Bonferroni cut-offs were applied to the L-A-LOR (ranging from 0.85 to 0.99), and Cox's B (ranging from 0.53 to 0.62) statistics to flag items with large DIF (also depending on the number of items per scale)¹.

The impact of DIF at scale level was examined by Differential Test Functioning (DTF) analyses. DTF concerns the additive effect of item DIF across all items of a scale. Penfield and Algina (2006) argued that a collective large level of DIF in a group of items exist if 25 % or more of the items show moderate to large DIF. They propose to define DIF effect variance (indicated by the weighted v^2 statistic) as small for $v^2 < .07$, medium for $.07 \leq v^2 \leq .14$, and large for $v^2 > .14$. In line with the Bonferroni corrected DIF thresholds, similar adjustments were made for the DTF

¹ The exact cutoff values for each of the three DIF indicators for each analysis can be obtained on requested from the first author.

thresholds². Both DIF and DTF analyses were conducted at facet level, so the total facet scores were used as stratifying variables. The stratum size was set at 1, which is the default option in DIFAS 5.0.

Results

Descriptive Statistics

Overall, for 16 of the 25 facets, the mean scale scores of the older sample differed significantly ($p \leq .05$) from the mean scale scores of the younger sample (see Table 1). Large significant differences ($d > .80$) were found for Intimacy Avoidance and Risk Taking, and moderate significant differences ($d > .50$) for Hostility, Attention Seeking, Deceitfulness, Manipulativeness, Distractibility, and Irresponsibility. For all but the Intimacy Avoidance scale, mean scale scores for these facets were significantly higher for younger compared to older adults ($p < .001$).

Differential Item and Test Functioning

Twenty-five separate DIF and DTF analyses were conducted for each of the primary traits of the PID-5, whereby the respective total facet-score served as stratifying variable. Overall, DIF analyses revealed 30 items showing significant DIF, divided across 15 facets. Table 2 shows the paraphrased item content and statistical details of these items, and the facets they belong to. In order to evaluate the impact of these DIF items at scale (i.e., facet) level, additional DTF analyses were conducted. According to the Bonferroni corrected cut-off (> 0.35)², DTF analyses revealed large DTF for Withdrawal, Attention Seeking, Rigid Perfectionism, and Unusual Beliefs (with weighted v^2 values of 0.64, 0.65, 0.37, and 0.44, respectively). Withdrawal contained five items displaying significant DIF. Two of them displayed DIF against the older age group (“I keep to myself” and “I keep my distance from people”), indicating they were more readily endorsed by younger adults with the same level of underlying personality

trait (i.e., Withdrawal). Three items showed DIF against younger adults (“I don’t like spending time with others”, “I’m not interested in making friends”, and “I say as little as possible when dealing with people”). Attention Seeking contained four items displaying significant DIF, of which two displayed DIF against younger (“I do things so that people just have to admire me”, and “I crave attention”), and two against older adults (“I love getting attention”, and “I like standing out in a crowd”). Rigid Perfectionism contained one item displaying DIF against older (“I simply won’t put up with things being out of their proper places”), and one item displaying DIF against younger adults (“I focus too much on minor details”). Similarly, Unusual Beliefs contained one item displaying DIF against older (“I believe that some people can move things with their minds”), and two displaying DIF against younger adults (“Other people seem to think my behavior is weird”, and “I see unusual connections between things”), indicating that the latter two are more readily endorsed by younger adults, matched on underlying personality trait.

Discussion

The primary aim of this study was to investigate the age-neutrality of the PID-5 facets by examining potential DIF for older versus younger adults. According to the stringent Bonferroni corrected cutoffs, analyses revealed a total of 33 items displaying significant DIF, divided across 15 facets. The impact hereof at scale level was relatively small. Large DTF was confirmed for four facets, namely Withdrawal, Attention Seeking, Rigid Perfectionism and Unusual Beliefs.

Withdrawal Three items showed negative DIF, indicating they were more readily endorsed by older adults with similar levels of the latent personality trait. These items focus on (the absence of) close relationships (e.g., ‘I’m not interested in making friends’). In this respect, it is possible that endorsing these items does not reflect personality pathology, but rather dealing with the death of loved ones, or, for example, isolation caused by physical illness (Van Alphen et al. 2006). Also, as people age, they tend to engage in selective social interaction, maintaining only the most rewarding contacts to satisfy their emotional needs (Carstensen 1991). Notably, two other items from the Withdrawal facet were more readily endorsed by younger adults (“I keep to myself” and “I keep my distance from people”). Although at first sight substantially very similar to the items displaying negative DIF, these findings might reveal a difference in the interpretation of these items that varies with age. It is not unlikely that older adults are less inclined to endorse these items that probe social isolation and withdrawal, because the diminishing of a social network is more common in later life, due to loss experiences or physical deterioration. Both younger and older participants

² In deriving their thresholds Penfield and Algina (2006) argued that a collective large level of DIF in a group of items exist if 25 % or more of the items are categorized as having moderate or large magnitudes of DIF based on the ETS classification scheme (i.e. if 25 % or more of the items have an absolute value of $\log(\alpha_{MH})$ greater than or equal to .43. They also suggest that MH and L-A LOR have similar meanings in terms of DIF magnitude. Because we wanted to reduce the Type I error, a Bonferroni correction was applied and adjusted L-A LOR cut-off values were used to flag items with large DIF (depending on the number of items per scale). In line, we made a similar adjustment for the DTF thresholds. For example: Penfield and Algina consider the variance of DIF effect large when weighted $v^2 > .14$, using an L-A LOR value of .43 as critical value. Since we adhere to stringent L-A LOR critical values (for example $> .92$ for Anhedonia) we adjusted this to $v^2 > .35$ (i.e., $.14/.43 \times .92$).

Table 2 Items meeting Bonferroni adjusted criteria for large DIF

Facet	Item	Paraphrased context	Mantel χ^2	L-A LOR	Cox's B
Items displaying DIF against younger adults (older > younger)					
Anxiousness	110.	I worry about almost everything.	9.72	-.69	-.61
Anxiousness	174.	I'm fearful about bad things that might happen.	10.98	-.71	-.58
Emotional lability	165.	I get emotional over every little thing.	18.54	-1.01	-.79
Restricted affectivity	167.	I never show emotions to others.	11.9	-.72	-.60
Withdrawal	136.	I don't like spending time with others.	12.58	-.98	-.78
Withdrawal	146.	I'm not interested in making friends.	17.42	-1.16	-.66
Withdrawal	147.	I say as little as possible when dealing with people.	27.64	-1.20	-.95
Attention seeking	43.	I do things so that people just have to admire me.	24.61	-1.24	-.84
Attention seeking	191.	I crave attention.	35.56	-1.36	-1.03
Callousness	207.	I don't see the point in feeling guilty about things.	17.29	-1.23	-.94
Irresponsibility	201.	I skip appointments if I'm not in the mood.	10.63	-.84	-.62
Rigid perfectionism	196.	I simply won't put up with things being out of their proper places.	37.40	-1.35	-.85
Risk taking	195.	I don't think about getting hurt when I'm doing things that might be dangerous.	16.36	-.98	-.66
Unusual beliefs	143.	People can move things with their minds.	12.60	-.99	-.68
Items displaying DIF against older adults (younger > older)					
Hostility	28.	I snap at people when they irritate me.	12.58	.78	.67
Hostility	32.	I can be mean when I need to be.	32.11	1.20	.77
Intimacy avoidance	89.	I keep romance out of my life.	11.46	1.02	.71
Withdrawal	20.	I keep to myself.	14.40	.87	.63
Withdrawal	82.	I keep my distance from people.	39.40	1.61	1.13
Attention seeking	74.	I love getting attention.	12.74	.79	.60
Attention seeking	111.	I like standing out in a crowd.	18.93	.97	.76
Emotional lability	18.	My emotions change for no good reason.	24.86	1.04	.74
Callousness	200.	I enjoy making people in control look stupid.	12.12	.86	.64
Deceitfulness	134.	I don't hesitate to cheat if it gets me ahead.	10.37	.79	.64
Deceitfulness	214.	Lying comes easily to me.	10.60	.77	.61
Manipulativeness	125.	Sweet-talking others helps me get what I want.	8.91	.68	.62
Distractibility	132.	I am easily distracted.	15.18	.92	.78
Rigid perfectionism	49.	I focus too much on minor details.	61.62	1.74	1.03
Eccentricity	24.	Others think my behavior is weird.	15.06	.93	.63
Unusual beliefs	194.	I see unusual connections between things.	14.31	.96	.67

L-A LOR Liu-Agresti common log odds ratio, Cox's B Cox's Noncentrality parameter estimator. ^R Reversed scored items. Facets with large DTF are given in bold

probably compared themselves to peers when filling out the PID-5 questionnaire, thus an item as “I keep to myself” may lead to less extremely high scores in an older population, because it is not so much seen as “deviant behavior”, but rather as a normative phenomenon related to aging.

Attention Seeking The item “I love getting attention” was more readily endorsed by younger adults, whereas the item “I crave attention” was more readily endorsed by older adults. Although the contents of these items show considerable overlap, it seems as though there really is a difference between “love attention” and “crave attention”. Craving attention might be a more desirable statement for those who are lonely, and, supposing elderly people might in

general be more lonely (e.g., Holmén and Furukawa 2002), this might explain the differences found for this item.

Rigid Perfectionism Younger adults scored generally higher on the item “I focus too much on minor details”. This item is possibly more related to an occupational context, and therefore less relevant (and less readily endorsed) by retirees. The other DIF item in this facet (“I simply won't put up with things being out of their proper places”) was more readily endorsed by older adults, suggesting that older adults might be in general more orderly and/or rigid, and less resistant against changes in their personal habitat than younger adults. Another explanation might lie in the cognitive decline and memory problems associated with advancing age. Older

adults are possibly more prone to compensate with order and regularity, as a way to cope with their forgetfulness.

Unusual Beliefs Two items displayed DIF, of which one was more readily endorsed by younger adults (“I see unusual connections between things”), and one by older adults (“People can move things with their minds”). The item “I see unusual connections between things” possibly reflects the tendency of young people to challenge traditional norms and their striving to be seen as unique individuals who want to gather information and expand their horizons, whereas older adults are less knowledge-oriented (Reed and Carstensen 2012).

Overall, the current initial results validate the comparison of mean facet scores across younger and older age groups for 21 of the 25 PID-5 traits. Given the lack of measurement invariance for Withdrawal, Attention Seeking, Rigid Perfectionism, and Unusual Beliefs, caution is warranted in the interpretation of age differences based on these particular traits. Concerning the current between-group differences on mean facet scores, largest mean differences (as evidenced by moderate and large effect sizes) were found for Hostility, Attention Seeking, Deceitfulness, and Manipulativeness, all four belonging to the higher-order trait domain of Antagonism. Younger adults had significantly higher mean level scores for these traits. Accordingly, younger adults scored also significantly higher on Risk Taking, Distractibility, and Irresponsibility, three facets of the Disinhibition domain. When subjected to a hierarchical structure analysis, it is demonstrated that Antagonism and Disinhibition are both split-offs of a higher-level Externalizing factor (Wright et al. 2012). These findings support the notion that externalizing personality traits tend to remit with age. Due to physical changes associated with aging and consequently a reduced mobility and slower pace, elderly people are less likely to act impulsively or manifest risky, irresponsible behavior (Roberts et al. 2006). Furthermore, the PID-5 five-factor structure shows clear resemblance to other five factor models like the Personality Psychopathology Five (PSY-5) domains, including Negative Emotionality/Neuroticism, Introversion/Low Positive Emotionality, Aggressiveness, Disconstraint, and Psychoticism (Anderson et al. 2013; Harkness and McNulty 1994), or the structure of normal personality as represented by the FFM, whereby Antagonism is the pathological variant of (low) Agreeableness, and Disinhibition the pathological variant of (low) Conscientiousness (Thomas et al. 2012). Parallels can thus be drawn between current findings and established research on age-related mean-level changes in general personality trait scores, describing an increase of Agreeableness and Conscientiousness traits throughout the lifespan (e.g., Terracciano et al. 2005; Roberts et al. 2006). Older adults, in turn, scored significantly higher on Intimacy Avoidance, a trait facet belonging to the higher-order domain Detachment, the pathological variant of (low) Extraversion. Again, these results corroborate the decline in Extraversion with advancing age.

Limitations

A few limitations should be considered. First of all, the cross-sectional design of the current study makes it difficult to distinguish real age effects from cohort effects, hampering the interpretation of the DIF results. It was not our intention however, to provide clear explanations for the current DIF results, but only to detect for possible DIF as a function of age in the recently proposed maladaptive personality traits for DSM-5. In line, the possibility cannot be ruled out that some of these initial findings are due to other factors beyond age that define differences between the current subsamples (e.g., gender, or socio-economic status). Future studies should therefore strive for larger, more representative samples and explore the possible underlying causes of DIF, and their consequences for the assessment of personality pathology across the lifespan in demographically matched samples of younger and older samples. The current sample sizes were too small to obtain stable parameter estimates using the graded response model or other polytomous IRT models, yet these more sophisticated research methods may play an important role in further exploring the possible influence of the above mentioned confounding variables [see e.g., Penfield (2010) for a detailed comparison between CTT and IRT approaches for identifying DIF]. Replicating the current findings within clinical samples is another important avenue for further research, since the PID-5 was primarily designed to identify personality pathology, a clinically-relevant phenomenon. A second limitation pertains to the Suspiciousness scale. Although no significant differences in mean scale scores were found between the younger and older age group for this trait, the internal consistency of this scale was clearly low in the current older sample and warrants further investigation. Possibly the two reverse scored items in this scale share variance above and beyond the general factor of suspiciousness, making them qualitatively distinct from the other items in the scale, and violating the assumption of local independence (e.g., Hopwood and Donnellan 2010). In addition, further research should explore whether a similar low internal consistency would appear in a clinical sample. Third, the current analyses and results apply to the Dutch version of the PID-5, so that further research is needed to investigate whether the same results would apply to, for instance, a North American population. Such cross-cultural research could possibly reveal interesting findings towards the universality of age-related changes in personality pathology and the course of personality disorders throughout the lifespan, aspects that are only just beginning to be understood. A last limitation considered here is the lower bound of age 60 as inclusion criteria for the older age group. The heterogeneity of this older age group should not be underestimated, since considerable differences might exist between, for example, 60–65 and 80–85 year-olds. With the current software used to investigate DIF we were limited to

the comparison of two age groups, but investigating measurement invariance across different age groups covering the whole lifespan and using smaller age-ranges might reveal interesting findings about more nuanced age-related changes in personality traits.

Conclusions

Despite not having explicitly considered the later life context during its development, current initial results show that most PID-5 traits are measured equally well across both a younger and an older age group. These results are promising in light of the growing awareness that an age-neutral measurement is crucial for a valid assessment of personality pathology throughout the lifespan. Additional research is certainly needed however to further refine this instrument and make it entirely age-neutral, since 33 items appeared to display large DIF, resulting in four scales exhibiting significant DTF. To this end, a set of alternative items could be written that works equally well for younger and older adults, regardless of their somewhat different living conditions. This set of items can then be tested for DIF across important demographic groups (e.g., gender, age, ethnic status, etc.), to finally reach a scale that contain no measurement artifacts (Oltmanns and Balsis 2011). We do realize that creating such items is a challenging task, however we hope that the current exploratory analyses point out the need and inspire further researchers towards developing an age-neutral measurement system.

References

- Agronin, M. E., & Maletta, G. (2000). Personality disorders in later life: Understanding and overcoming the gap in research. *The American Journal of Geriatric Psychiatry*, 8, 4–18.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Anderson, J. L., Sellbom, M., Bagby, M. R., Quilty, L. C., Veltri, C. O. C., Markon, K. E., & Krueger, R. F. (2013). On the convergence between PSY-5 domains and PID-5 domains and facets: implications for assessment of *DSM-5* personality traits. *Assessment*. doi:10.1177/1073191112471141.
- Baer, R. A., Samuel, D. B., & Lykins, E. L. B. (2011). Differential item functioning on the five facet mindfulness questionnaire is minimal in demographically matched mediators and nonmediators. *Assessment*, 18(1), 3–10.
- Balsis, S., Gleason, M. E. J., Woods, C. M., & Oltmanns, T. F. (2007). An item response theory analysis of DSM-IV personality disorder criteria across younger and older age groups. *Psychology and Aging*, 22(1), 171–185.
- Camilli, G., & Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. *Journal of Educational and Behavioral Statistics*, 24, 323–341.
- Carstensen, L. L. (1991). Selectivity theory: social activity in life-span context. *Annual Review of Gerontology and Geriatrics*, 11, 195–217.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- De Clercq, B., De Fruyt, F., Mervielde, I., Krueger, R. F., & Markon, K. (2011). Dutch translation of the Personality Inventory for DSM-5 (PID5).
- Edwards, M. C., & Edelen, M. O. (2009). Special topics in item response theory. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 178–185). London: Sage.
- Harkness, A. R., & McNulty, J. L. (1994). The personality psychopathology five (PSY-5): Issue from the pages of a diagnostic manual instead of a dictionary. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 291–315). New York: Springer.
- Holmén, K., & Furukawa, H. (2002). Loneliness, health and social network among elderly people—a follow-up study. *Archives of Gerontology and Geriatrics*, 35, 261–274.
- Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14(3), 332–346.
- Hopwood, C. J., Thomas, K. M., Markon, K. E., Wright, A. G. C., & Krueger, R. F. (2012). DSM-5 personality traits and DSM-IV personality disorders. *Journal of Abnormal Psychology*, 121(2), 424–432.
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, 42, 1879–1890.
- Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratio with a stratified ordinal response. *Biometrics*, 52, 1223–1234.
- Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, 52, 443–451.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Oltmanns, T. F., & Balsis, S. (2011). Personality disorders in later life: questions about the measurement, course, and impact of disorders. *Annual Review of Clinical Psychology*, 7, 321–349.
- Penfield, R. D. (2005). DIFAS: differential item functioning analysis system. *Applied Psychological Measurement*, 29, 150–151.
- Penfield, R. D. (2007). DIFAS 4.0. Differential item functioning analysis system: user's manual. Unpublished manuscript.
- Penfield, R. D. (2010). Distinguishing between net and global DIF in polytomous items. *Journal of Educational Measurement*, 47, 129–149.
- Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement*, 43(4), 295–312.
- Reed, A. E., & Carstensen, L. L. (2012). The theory behind the age-related positivity effect. *Frontiers in Psychology*, 3, doi:10.3389/fpsyg.2012.00339.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin*, 132, 3–27.
- Segal, D. L., Coolidge, F. L., & Rosowsky, E. (2000). Personality disorders. In S. Krauss (Ed.), *Psychopathology in later adulthood* (pp. 89–115). Wiley: New York.
- Tackett, J. L., Balsis, S., Oltmanns, T. F., & Krueger, R. F. (2009). A unifying perspective on personality pathology across the life span:

- developmental considerations for the fifth edition of the diagnostic and statistical manual of mental disorders. *Development and Psychopathology*, 21, 687–713.
- Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of NEO PI-R scales in the Baltimore longitudinal study of aging. *Psychology and Aging*, 20(3), 493–506.
- Thomas, K. M., Yalch, M. M., Krueger, R. F., Wright, A. G. C., Markon, K. E., & Hopwood, C. J. (2012). The convergent structure of DSM-5 personality trait facets and Five Factor Model trait domains. *Assessment*.
- Van Alphen, S. P. J., Engelen, G. J. J. A., Kuin, Y., & Derksen, J. J. L. (2006). The relevance of a geriatric sub-classification of personality disorders in the DSM-V. *International Journal of Geriatric Psychiatry*, 21, 205–209.
- Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: shifting to a dimensional model. *American Psychologist*, 62, 71–83.
- Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology*. doi:[10.1037/a0027669](https://doi.org/10.1037/a0027669).